



# In Fisher's net : exact F-tests in semi-parametric models with exchangeable errors

Frédéric Jouneau-Sion, Olivier Torrès

## ► To cite this version:

Frédéric Jouneau-Sion, Olivier Torrès. In Fisher's net : exact F-tests in semi-parametric models with exchangeable errors. 2014. halshs-01062623

**HAL Id: halshs-01062623**

**<https://shs.hal.science/halshs-01062623>**

Preprint submitted on 10 Sep 2014

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

WP 1422

**In Fisher's net : exact F-tests in semi-parametric  
models with exchangeable errors**

Frédéric Jouneau-Sion, Olivier Torrès

September 2014

**GATE Groupe d'Analyse et de Théorie Économique Lyon-St Étienne**

93, chemin des Mouilles 69130 Ecully – France

Tel. +33 (0)4 72 86 60 60

Fax +33 (0)4 72 86 60 90

6, rue Basse des Rives 42023 Saint-Etienne cedex 02 – France

Tel. +33 (0)4 77 42 19 60

Fax. +33 (0)4 77 42 19 50

Messagerie électronique / Email : [gate@gate.cnrs.fr](mailto:gate@gate.cnrs.fr)

Téléchargement / Download : <http://www.gate.cnrs.fr> – Publications / Working Papers

# In Fisher's net : exact F-tests in semi-parametric models with exchangeable errors <sup>1</sup>

Jouneau-Sion Frédéric <sup>2, 3</sup>  
Torrès Olivier <sup>4</sup>

## abstract

We consider testing about the slope parameter  $\beta$  when  $Y - X\beta$  is assumed to be an exchangeable process conditionally on  $X$ . This framework encompasses the semi-parametric linear regression model. We show that the usual Fisher's procedure have non trivial exact rejection bound under the null hypothesis  $R\beta = \gamma$ . This bound derives from the Markov inequality and a close inspection of multivariate moments of self-normalized, self-centered, exchangeable processes. Improvement by higher order versions of the Markov inequality are also presented. The bounds do not require the existence of any moment, so they remain valid even if TCL do not apply. We generalize the framework to multivariate and order-1 auto-regressive models with exogenous variables.

JEL : C01, C12, C14.

## 1 Introduction

Despite widespread usage of asymptotic methods in econometric, exact techniques are important tools for several reasons. First, consider for instance the increasingly important field of spatially indexed data-set. Changing the geographical scale allows to enlarge the sample size, but it also implies institutional changes since countries, states, counties serves different needs. In strategic contexts, increasing the sample size even by one unit may change the whole outcome. Indeed one player more and/or one more step typically change the behaviors and the set of Nash equilibria. <sup>5</sup> As asymptotic approximations rely on an sequence of increasingly large samples, they are difficult to use in such contexts.

Second, asymptotic approximations often rely on point-wise convergence results. Asymptotic control of type-one error is not insured when nuisance parameters are present under the null hypothesis. The literature on poor asymptotic approximations is now very large : see, among others, Nankervis and N.E. Savin [1985] or Mikusheva [2007] in the context of dynamic models, Pötscher [2009] on penalized estimators, Dufour [1997,2003] on weak instruments.

Finally, most –if not all– asymptotic methods requires the existence of moments. <sup>6</sup> Now, recent studies in finance, insurance, income and wealth inequality

---

<sup>1</sup>We are indebted to Jean-Yves Welschinger for an important improvement of the paper. We also thank Lynda Khalaf and Philippe Barbe for valuable comments.

<sup>2</sup>Université de Lyon, Lyon, F-69007, France ; CNRS, GATE Lyon Saint-Etienne, Ecully, F-69130, France

<sup>3</sup>Université Lyon 2, Lyon, F-69007, France

<sup>4</sup>EQUIPPE-GREMARS Universités de Lille

<sup>5</sup>In the limit, an infinite number of players may even lead to complete disappearance of strategic behaviors. On the other side, it is a well documented fact that the set of Nash equilibria in infinitely repeated games is typically very large.

<sup>6</sup>Typically, squared integrability is required.

emphasize the role of Pareto-tail behavior for many important economic series (see Gabaix [2009] for an overview). The asymptotic behavior of usual statistics may be affected such Pareto-like tail and even Bootstrap approximation may fail in this case.

Several recent progresses have been made in exact prior-free inference techniques. Notably, simulation-based methods (see Dufour and Khalaf [2002] or Beaulieu, Dufour and Khalaf [2013] and Dufour [2006]) provide several ways to derive exact tests and/or confidence region in a very large class of models. In particular, the projection method makes it possible to overcome the difficult problem of nuisance parameters (see, for instance, Dufour and Taamouti [2005]). Recently, Gossner and Schalg [2013] propose exact approaches for the linear model in the case of a bounded dependent variable with known bounds.

One major limitation of many exact methods is that they require specific procedures. The purpose of the current paper is to provide exact prior-free control of type-I error of the Fisher test procedure. As it is well known, the Fisher test is one of the most common procedure in linear model. The main justification of this procedure rests on some invariance principles and on asymptotic efficiency. Yet, Fisher and Student tests are also known for possibly arbitrarily large deviation of their ‘claimed’ level when the distribution departs from the Gaussian assumption. In this paper, we provide explicit, non trivial, exact bounds for the size of a Fisher test.

In our framework, the models are semi-parametric *i.e.* contain a nuisance parameter taking value in a infinite dimensional space. No assumption about the median is necessary. Moreover the framework covers cases in which tails may be as fat as needed. In particular the existence of moments is not required. This implies that our approach may be used even when the particular assumptions needed for the CLT to hold fail to be met.

When no assumption but iidness is made on the error term of a linear model, an extension of a result by Bahadur and Savage [1953] shows that exact tests about the expectation are always trivial (see Dufour, Jouneau-Sion and Torres [2008]). ‘Trivial’ means here that the power of any test cannot exceed its size for any value of the parameter fixed under the null. Now, a linear model typically contains two types of parameters: the slopes and the intercept. Although BS’s result applies to the *entire* parameter, it does not necessarily preclude the existence of non trivial test for *parts* of the parameter. It turns out that non trivial tests about the vector of slopes do exist and that Fisher’s procedure is one of them.

The results rely on the following argument. When testing linear constraints on the slope parameters in a model containing an intercept parameter, the Fisher statistic depends under the null hypothesis on self-centered, self-normalized versions of the original vector of errors. A self-centered vector is obtained after difference with the in-sample average. A self-normalized vector is obtained after division by the in-sample standard error. Now, if the vector of error terms is an exchangeable process (and, in particular, if unobserved disturbances are iid) the Fisher statistic is driven under the null by a self-centered, self-normalized exchangeable process.<sup>7</sup> The first and second order multivariate moments of such

---

<sup>7</sup>The literature dealing with exchangeability often consider processes. In our setting, we consider finite sample vectors. As the expression ‘exchangeable random vector’ is uncommon with shall refer to ‘exchangeable process’ even though we do not consider increasingly large samples.

processes is shown to depend solely on the sample size. Applying the Markov inequality we deduce a distribution-free boundary for the rejection probability under the null. To the best of our knowledge, such a bound on the Fisher statistic is original.<sup>8</sup> Our results may be used to recompute exact distribution-free rejection probabilities when original data are no longer available but a tabulated Fisherian p-value is provided.<sup>9</sup>

As self-normalized processes are bounded, multivariate moments of any order exist, and higher order improvements of the Markov/Chebyshev inequality may be used. Such improvements raise specific difficulties. First, they cannot be strictly better since the boundary derived from Markov inequality is shown to be reached in some particular –albeit extreme– cases. Second, multivariate moments of a self-centered, self-normalized exchangeable process depend on nuisance parameters when the order is strictly larger than 2. Exact control of type-I error then require to maximize over all possible values of these moments. This control raises challenging questions that are addressed in details.

We propose two extensions to multivariate and dynamic models. The latter extension exploits an exact testing approach to AR(1) model with exogenous explanatory variable proposed by Dufour and Kiviet [1998]. Contrarily to Dufour and Kiviet [1998] our framework does not require knowledge of the distribution of shocks. This extension is computationally demanding as the statistic depends on eight-order multivariate moments. Multivariate models lead to consider a weaker form of exchangeability (namely line-exchangeability). This feature emerges since ‘blocks’ of observations coming from two different sample unit (say ‘individual’) may be considered as –conditionally– exchangeable, whereas exchangeability is likely to fail if we consider observations coming from the same individual.<sup>10</sup>

The paper is organized as follows. Section 2 presents the case of a univariate linear model in the context of cross-sectional data. Section 3 examines the extension to an ARX(1) model. We illustrate the technique in the particularly documented context of unit root testing in presence of a linear trend. Section 4 extends the approach to multivariate models.

## 2 Exact testing in a linear model defined by exchangeability of the error terms

In this section, we consider a model in which the difference between a random vector  $Y$  playing the role of the endogenous variable and a given linear combination of “explanatory variables” represented by some random matrix  $X$  is exchangeable conditionally on  $X$ .

We first set up the model and the parameters. Then we derive an exact bound for the Fisher test about this parameter. We show that this bound

---

<sup>8</sup>The fact that self-normalized, self-centered exchangeable processes have known multivariate order-1 and order-2 moments has been exploited by Dufour and Roy [1985] in the context of AR models.

<sup>9</sup>A vast body of literature deals with power of the Fisher testing procedure. In particular the consistency of this test is now established under fairly general conditions.

<sup>10</sup>This weaker form of exchangeability could also be considered to handle cases in which data are affected by some given sort of heteroskedasticity. We shall not pursue this idea further in the current paper.

is ‘tight’ by providing an example in which it is reached. We finally discuss improvements of the initial bound.

## 2.1 A linear model with exchangeable error terms

We consider a family  $\mathcal{P}$  of probability distributions for  $(Y, X)$ , where  $Y, X$  are matrices of real random variables of size  $n \times 1$  and  $n \times p$  respectively (with  $p < n$ ). In this section, we assume the following condition holds

**Assumption 1** *For any  $\mathbf{P} \in \mathcal{P}$  there exist a single size- $p$  vector denoted  $\beta(\mathbf{P})$  such that the distribution of  $Y - X\beta(\mathbf{P})$  conditional on  $X$  is a non degenerated exchangeable process.*

Exchangeable processes have a long history in statistics that goes back at least to de Finetti theorem in 1931 and several seminal papers by Aldous in the 80’s (see, for an overview, Aldous [1983]). A classical references are Galambos [1982] or Kendall, Stuard and Ord [1983]. The usefulness of this assumption in econometrics is rarely mentioned one exception being McCullagh [2005]. In particular McCullagh [2005] emphasizes and clarifies the link between exchangeability, modelization and measurment of treatment effects. Dufour and Roy [1985] exploits exchangeability to test for autocorrelation.

It is of crucial importance to notice that the uniqueness of  $\beta(\mathbf{P})$  in the above condition implies that the vector space spanned by the realizations of the columns of  $X$  does not contain  $\iota_n = (1, \dots, 1)' \in \mathbb{R}^n$  a.s. Indeed, if this not the case, we may find  $v \neq 0$  such that  $Xv = \iota_n$ . Thus  $Y - X\beta(\mathbf{P})$  and  $Y - X(\beta(\mathbf{P}) + v)$  are both exchangeable conditionally on  $X$ .

Notice that uniqueness of  $\beta(\mathbf{P})$  in Assumption 1 also implies the usual identification condition. If  $\text{rank}(X) < p$ , we may find  $u \neq 0$  such that  $Xu = 0$ , which would make  $Y - X(\beta(\mathbf{P}) + v)$  exchangeable conditionnally on  $X$ .

A few word are needed to interpret the parameter  $\beta$ . For any size- $n$  vector  $Z$  let  $Z_i$  denotes the  $i$ -th component of this vector. Assume for any  $i \in \{1, \dots, n\}$  an any (measurable) event  $A$  we have

$$\mathbf{P}((Y - X(\beta(\mathbf{P})))_i \in A|X) = \mathbf{P}((Y - X(\beta(\mathbf{P})))_i \in A|X_i)$$

This assumption is common in ‘cross section’ settings.

As exchangeability implies identical distributions of the components, we have for all  $i, j$

$$\begin{aligned} \mathbf{P}((Y - X(\beta(\mathbf{P})))_i \in A|X) &= \mathbf{P}((Y - X(\beta(\mathbf{P})))_i \in A|X_i) \\ \mathbf{P}((Y - X(\beta(\mathbf{P})))_j \in A|X) &= \mathbf{P}((Y - X(\beta(\mathbf{P})))_i \in A|X) \\ \mathbf{P}((Y - X(\beta(\mathbf{P})))_i \in A|X) &= \mathbf{P}((Y - X(\beta(\mathbf{P})))_j \in A|X) \end{aligned}$$

Therefore

$$\mathbf{P}((Y - X(\beta(\mathbf{P})))_i \in A|X_i) = \mathbf{P}((Y - X(\beta(\mathbf{P})))_j \in A|X_j) \forall i, j \in \{1, \dots, n\}^2$$

In words, the components of  $Y - X\beta(\mathbf{P})$  are identically distributed conditionally on  $X$  and are independent from  $X$ . In this case, the distribution of  $Y_i$  of the vector  $Y$  depends on  $X$  only through  $(X\beta(\mathbf{P}))_i$ . For instance a modification ”everything else equal” of the  $i$ -th line of matrix  $X$  affect the quantiles of  $Y_i$

through  $(X\beta(\mathbf{P}))_i$ . In case of existence this also hold true for the expectation, but we stress that the existence of any moment of  $Y$  is not required for our results to hold.<sup>11</sup>

The above interpretation makes clear that, the parameter of interest  $\beta$  contains the “slopes” of some linear dependence of  $Y$  on  $X$ . Notice however  $\beta$  *cannot* contain an intercept. In a sense, the intercept term may be viewed as part of the “disturbances”  $Y - X\beta(\mathbf{P})$  since if a size- $n$  random vector  $v$  is exchangeable then so is  $a\iota_n + v$ .

In many applications, the intercept is a nuisance parameter, yet there is a deeper reason to avoid the intercept being part of  $\beta$ . Dufour, Jouneau-Sion and Torrès [2008] establish that meaningful valid test about the intercept do not exist even under the stronger condition that the vector  $Y - X\beta(\mathbf{P})$  is iid conditionally on  $X$ . More precisely, consider any null hypothesis that fixes the value of the intercept parameter, their results implies that any test at level  $\alpha \in [0, 1]$  must have power  $\alpha$  whatever the sample size.<sup>12</sup>

It is also important to stress that the exchangeability hypothesis concerns the conditional distribution of the “disturbances” and not the process  $(Y, X)$  itself. We do not impose  $(Y, X)$  iid. Dufour, Jouneau-Sion and Torrès [2008] show that if  $(Y, X)$  were iid, for any hypothesis on  $\beta$ , the power of any test with level  $\alpha$  cannot exceed  $\alpha$ .

Finally, the non degeneracy assumption is made to insure the existence of the Fisher statistic.

## 2.2 Markov bound for the type-I error in the Fisher test

In what follows,  $\bar{\mathbf{P}}$  denotes the distribution of  $(Y, X)$  that generated the available data set.

Let  $R$  be a  $q \times p$  real matrix with rank  $q$ . Consider the null hypothesis  $H_0 : R\beta(\mathbf{P}) = \gamma_0$ . We write  $\mathbf{P} \in \mathcal{P}_0$  whenever  $\mathbf{P}$  satisfies Assumption 1 and  $R\beta(\mathbf{P}) = \gamma_0$ .

The Fisher test statistic associated with  $H_0$  is

$$F = \nu \frac{(R\hat{\beta} - \gamma_0)'(Q(Z'Z)^{-1}Q')^{-1}(R\hat{\beta} - \gamma_0)}{\hat{\varepsilon}'\hat{\varepsilon}}$$

where  $Z \equiv (\iota_n, X)$ ,  $Q \equiv (0, R)$ ,  $\nu \equiv (n - p - 1)/q$ ,  $\hat{\theta} = (\hat{\mu}, \hat{\beta})'$  is the OLS estimator associated with the linear regression of  $Y$  on  $Z$  and  $\hat{\varepsilon}$  is the corresponding vector of OLS residuals, respectively.

It is important to stress that although the null hypothesis does not fix the value of the intercept term, the results we are about to present concern models in which the intercept is a nuisance parameter. From a practical viewpoint, it

<sup>11</sup>Remark also that the ‘cross section’ assumption we used to *interpret*  $\beta$  is not required to derive the results presented below.

<sup>12</sup>This problem arises even if the existence of moments of any order is assumed. It appears because *uniform* control of type-I error over all possible values of the ‘non parametric’ nuisance parameter (namely the distribution of the error terms) induces an arbitrary large loss of power whatever the fixed value of the parametric part of the model. On the other hand, test procedures whose power against *some* alternative is strictly larger than  $\alpha \in ]0, 1[$  do not fulfill the level constraint at level  $\alpha$ . Ultimately, consistent tests must have a level equal to 1 in the limit. Such dramatic loss of control of type-I error are now well documented in several contexts see *e.g.* Pötscher [2009], Dufour [1997, 2003] or Mikusheva [2007].



means the regression *must* be ran with an intercept parameter for our results to hold.

**Proposition 1** *For all  $c > 0$  we have*

$$\sup_{\mathbf{P} \in \mathcal{P}_0} \mathbf{P}(F \geq c | X) \leq \frac{n-p-1}{c(n-1)} + \frac{p}{n-1} \quad (1)$$

*Proof*: Pick a  $\mathbf{P}$  in  $\mathcal{P}_0$ . Partitioning  $Z'Z$  as

$$Z'Z = \begin{pmatrix} \iota'_n \iota_n & \iota'_n X \\ X' \iota_n & X'X \end{pmatrix}$$

and using block inversion, we get that the corresponding lower-right block of  $(Z'Z)^{-1}$  is  $(\tilde{X}'\tilde{X})^{-1}$ , where  $\tilde{X}$  is the projection of  $X$  on the orthogonal of the vector space spanned by  $\iota_n$ , i.e.,  $\tilde{X} \equiv M_{\iota_n}X$  with  $M_{\iota_n} \equiv I_n - \iota_n \iota'_n / n$ . Therefore we have  $Q(Z'Z)^{-1}Q' = R(\tilde{X}'\tilde{X})^{-1}R'$  and the statistic may be written as

$$F = \nu \frac{(R\hat{\beta} - \gamma_0)'(R(\tilde{X}'\tilde{X})^{-1}R')^{-1}(R\hat{\beta} - \gamma_0)}{\hat{\varepsilon}'\hat{\varepsilon}} \quad (2)$$

Next notice the well known Frish-Waugh theorem establishes that  $\hat{\beta}$  and  $\hat{\varepsilon}$  may be obtained from the OLS applied to

$$\tilde{Y} = \tilde{X}\beta + u$$

where  $\tilde{Y} \equiv M_{\iota_n}Y$  and  $u \equiv M_{\iota_n}(Y - X\beta(\bar{\mathbf{P}}))$ . Formally,  $\hat{\beta} = (\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{Y}$  and  $\hat{u} \equiv \tilde{Y} - \tilde{X}\hat{\beta} = \hat{\varepsilon}$ . Therefore,  $\hat{\beta} = \beta(\bar{\mathbf{P}}) + (\tilde{X}'\tilde{X})^{-1}\tilde{X}'u$  and  $\hat{\varepsilon} = M_{\tilde{X}}u$  where  $M_{\tilde{X}} \equiv I_n - \tilde{X}(\tilde{X}'\tilde{X})^{-1}\tilde{X}'$ . Under  $H_0$  we may write  $R\hat{\beta} - \gamma_0 = R(\tilde{X}'\tilde{X})^{-1}\tilde{X}'u$  and from the expression (2), the distribution of  $F$  under  $\mathbf{P} \in \mathcal{P}_0$  conditional on  $X$  is seen to be the same as that of

$$\nu \frac{u'\tilde{X}\Lambda\tilde{X}'u}{u'M_{\tilde{X}}u}$$

where  $\Lambda = (\tilde{X}'\tilde{X})^{-1}R'[R(\tilde{X}'\tilde{X})^{-1}R']^{-1}R(\tilde{X}'\tilde{X})^{-1}$ .

The Fisher test procedure rejects the null whenever, for a given real number  $c$ , we observed  $F \geq c$ . Under any  $\mathbf{P} \in \mathcal{P}_0$ , the probability of this event conditional on  $X$  is the same as the conditional probability of observing  $u'\tilde{X}\Lambda\tilde{X}'u \geq \kappa u'M_{\tilde{X}}u$ , where  $\kappa = c/\nu$ . Using the definition of  $M_{\tilde{X}}$ , straightforward matrix algebra shows this inequality is equivalent to  $\eta'B\eta \geq \kappa$ , where  $\eta = u/\sqrt{u'u}$  and  $B = \tilde{X}[\Lambda + \kappa(\tilde{X}'\tilde{X})^{-1}]\tilde{X}'$ . Notice that the process  $\eta_1, \dots, \eta_n$  is exchangeable conditionally on  $X$ . Indeed, from the definition of  $u$  and  $\eta$  we have

$$\eta_i = \frac{\tilde{Y}_i - (\tilde{X}\beta(\bar{\mathbf{P}}))_i}{\sqrt{\sum_{i=1}^n (\tilde{Y}_i - (\tilde{X}\beta(\bar{\mathbf{P}}))_i)^2}}, \quad i = 1, \dots, n$$

Moreover  $\sum_{i=1}^n \eta_i = 0$  and  $\sum_{i=1}^n \eta_i^2 = 1$ . Using exchangeability and these constraints, we get for all  $i, j = 1, \dots, n$ ,

$$\mathbf{E}_{\mathbf{P}}(\eta_i \eta_j | X) = \begin{cases} \frac{1}{n} & \text{if } i = j \\ \frac{-1}{n(n-1)} & \text{otherwise} \end{cases}$$

Now  $E_{\mathbf{P}}(\eta' B \eta | X) = \text{tr}[B E_{\mathbf{P}}(\eta \eta' | X)]$  where  $\text{tr}$  denotes the trace operator. We have  $E_{\mathbf{P}}(\eta \eta' | X) = \frac{1}{n} I_n - \frac{1}{n(n-1)} (\iota_n \iota_n' - I_n) = \frac{1}{n-1} I_n - \frac{1}{n(n-1)} \iota_n \iota_n'$ . From this and  $B \iota_n = 0$  it follows  $B E_{\mathbf{P}}[\eta \eta' | X] = \frac{1}{n-1} B$ , and thus

$$E_{\mathbf{P}}(\eta' B \eta | X) = \frac{1}{n-1} \text{tr}(B)$$

Now

$$\text{tr}(B) = \text{tr}(\tilde{X} \Lambda \tilde{X}') + \kappa \text{tr}(\tilde{X}(\tilde{X}' \tilde{X})^{-1} \tilde{X}')$$

Under the assumptions made on the ranks of  $X$  and  $R$  we have  $\text{tr}(B) = q + \kappa p$ . We then obtain

$$E_{\mathbf{P}}(\eta' B \eta | X) = \frac{1}{n-1} (q + \kappa p).$$

As  $B$  is a definite positive matrix, the Markov inequality then yields

$$\mathbf{P}(F \geq c | X) = \mathbf{P}(\eta' B \eta \geq \kappa | X) \leq \frac{1}{\kappa(n-1)} (q + \kappa p)$$

Finally, using the definition of  $\kappa$  we have

$$\mathbf{P}(F \geq c | X) \leq \frac{n-p-1}{c(n-1)} + \frac{p}{n-1}$$

This is true for any choice of  $\mathbf{P} \in \mathcal{P}_0$ .

*Q.E.D*

### 2.3 Universal corrections on p-values

The Markov bound derived in Proposition 1 allows for a uniform control of the type-I error of the test. For instance, a test of  $H_0$  at level  $\alpha$  requires choosing  $c$  so that the LHS in (1) is equal to  $\alpha$ . When  $p = 1$  (a regression model with a single explanatory variable) we must set  $c = \frac{n-2}{n\alpha - (\alpha+1)}$ , provided  $n$  is large enough.<sup>13</sup> In a bilateral test, we then reject  $H_0$  at level  $\alpha$  when the absolute value of usual Student  $t$  statistic associated with  $H_0$  is larger than  $\sqrt{\frac{n-2}{n\alpha - (\alpha+1)}}$ . In large sample, this critical value is equivalent to  $1/\sqrt{\alpha}$ . In such a case, when we choose  $\alpha = 5\%$ , the usual 1.96 limit (gaussian) critical value should be replaced by  $\sqrt{20} \simeq 4.47$ .

It is remarkable to note that the above bound solely depends on the model tested and the sample size. As a consequence, we may recompute a semi-parametric upper bound for the level of a test derived under the usual asymptotic assumption. For instance, if  $n$  is large wrt  $p$  the upper bound is close to  $1/c$ . The following table provides the inverse of the critical points of the Fisher when the sample size is very large as a function of the level and the number of constraints.

---

<sup>13</sup>Recall we must have  $c > 0$ .

Table 1: p-value correspondences in large samples

	Number of constraints									
	1	2	3	4	5	6	7	8	9	10
level										
$10^{-6}$	0.04	0.07	0.10	0.12	0.13	0.15	0.17	0.18	0.20	0.21
$10^{-5}$	0.05	0.09	0.11	0.14	0.16	0.18	0.20	0.21	0.23	0.24
$10^{-4}$	0.07	0.11	0.14	0.17	0.19	0.21	0.23	0.25	0.26	0.28
0.001	0.09	0.14	0.18	0.21	0.24	0.26	0.29	0.3	0.32	0.33
0.0025	0.11	0.17	0.21	0.24	0.27	0.29	0.31	0.33	0.35	0.37
0.01	0.15	0.22	0.26	0.30	0.33	0.35	0.38	0.40	0.41	0.43
0.025	0.20	0.27	0.32	0.36	0.39	0.41	0.43	0.45	0.47	0.49
0.05	0.26	0.33	0.38	0.42	0.45	0.47	0.5	0.51	0.53	0.54
0.10	0.37	0.43	0.48	0.51	0.54	0.56	0.58	0.60	0.61	0.62
0.15	0.48	0.53	0.56	0.59	0.61	0.63	0.65	0.66	0.68	0.69
0.20	0.61	0.62	0.65	0.67	0.68	0.70	0.71	0.72	0.73	0.74
0.25	0.75	0.72	0.73	0.74	0.75	0.76	0.77	0.78	0.79	0.8

The above table should be read as follows. If the level chosen is 5% and the null hypothesis concerns one constraint, we know for sure that for any distribution of the error term and any distribution of the explanatory variables, the p-value cannot be larger than 25%. Also, if we want to be sure that the p-value is smaller than 5% we need the p-value of the usual asymptotic Fisher test to be smaller than  $10^{-4}$ . Recall this correction holds even if no moment conditions are assumed. In particular they can be used when necessary conditions for the central limit theorem to apply do not hold.<sup>14</sup>

## 2.4 Tightness of the Markov bound

We now show that the bound derived in proposition 1 may be reached (at least for a specific distribution of the explanatory variables  $X$ .) In what follows we use the notation introduced in the proof of proposition 1 and  $[\cdot]$  is the integer part function on  $\mathbb{N}$ .

**Proposition 2** *For all  $n \geq 2$  and for all  $c$  in*

$$\left\{ \frac{n(n-2)}{k(n-1)-n} \mid k \in \mathbb{N}, 2 \leq k \leq \lfloor n/2 \rfloor; k/2 = \lfloor k/2 \rfloor \right\}$$

*there exist  $\mathbf{P} \in \mathcal{P}_0$  such that*

$$\mathbf{P}(F \geq c \mid X) = \frac{n-p-1}{c(n-1)} + \frac{p}{n-1} \quad (3)$$

*Proof:* Let  $n$  be an integer strictly larger than 4 and consider  $\varepsilon_i = \mathbb{1}(i = \nu)$ ,  $i = 1, \dots, n$ , where  $\nu$  is a uniform random variable on  $\{1, 2, \dots, n\}$ . It is clear that  $\varepsilon_1, \dots, \varepsilon_n$  forms an exchangeable sequence.

<sup>14</sup>For such necessary and sufficient conditions see, *e.g.* Dedecker and Merlevède [2002].

Easy computation leads to

$$\eta_i = \varepsilon_i \sqrt{1 - 1/n} - (1 - \varepsilon_i) \sqrt{1/[n(n-1)]}.$$

Consider the case with a single explanatory variable and assume  $\gamma_0 = 0$ . Also assume the the marginal distribution of  $X$  is such that  $\sum_{i=1}^n X_i = 0$  and  $\sum_{i=1}^n X_i^2 = 1$ . We have  $X = \tilde{X} = (X_1, \dots, X_n)'$  so that  $\Lambda = (X'X)^{-1} = 1$ . Thus  $\eta' B \eta = (\eta' X)^2 (1 + \kappa)$ . Now  $\kappa = c/(n-2)$  and then

$$\eta' B \eta \geq \kappa \iff \left( \sum_{i=1}^n \eta_i X_i \right)^2 \geq \frac{c}{c+n-2}$$

As  $\sum_{i=1}^n X_i = 0$ , we also derive from the definition of  $\varepsilon_i$  that for any  $i_0 = 1, \dots, n$

$$\mathbf{P} \left( \sum_{i=1}^n \eta_i X_i = \sqrt{\frac{n}{n-1}} X_{i_0} \middle| X \right) = \frac{1}{n}$$

From this result we deduce

$$\mathbf{P}(\eta' B \eta \geq \kappa | X) = \mathbf{P} \left( \left( \sum_{i=1}^n \eta_i X_i \right)^2 \geq \frac{c}{c+n-2} \middle| X \right) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}(X_i^2 \geq \frac{c}{c+n-2} \frac{n-1}{n})$$

To show that our bound is tight we need to exhibit a sequence  $X_1, \dots, X_n$  such that  $\sum_{i=1}^n X_i = 0$ ,  $\sum_{i=1}^n X_i^2 = 1$  and

$$\frac{1}{n} \sum_{i=1}^n \mathbb{1}(X_i^2 \geq \frac{c}{c+n-2} \frac{n-1}{n}) = \frac{c+n-2}{c(n-1)}$$

Now as the left hand side equals  $k/n$  for some  $k = 0, \dots, n$  so must be the right hand side. Moreover, we know that  $c > 0$  hence  $(c+n-2)/(c(n-1)) > 1/(n-1)$  so that  $k > 1$ . Finally if we set  $\frac{k}{n} = \frac{c+n-2}{c(n-1)}$  we have

$$\frac{c}{c+n-2} \frac{n-1}{n} = \frac{1}{k}$$

For any even integer  $k = 2k' > 1$  such that  $k < n/2$  define the following real sequence

$$X_i = \frac{\mathbb{1}(1 \leq i \leq k')}{\sqrt{k}} - \frac{\mathbb{1}(k'+1 \leq i \leq k)}{\sqrt{k}}$$

It is clear  $\sum_{i=1}^n X_i = 0$  and  $\sum_{i=1}^n X_i^2 = 1$ . Moreover we have

$$\sum_{i=1}^n \mathbb{1}(X_i^2 \geq \frac{c}{c+n-2} \frac{n-1}{n}) = \sum_{i=1}^n \mathbb{1}(X_i^2 \geq \frac{1}{k}) = k$$

Finally the condition  $\frac{c}{c+n-2} \frac{n-1}{n} = \frac{1}{k}$  implies

$$c = \frac{n(n-2)}{k(n-1) - n}$$

and notice that  $k > 1$  implies  $k(n-1) - n > 0$  hence  $c > 0$ .

*Q.E.D*

In the example provided in the proof we have  $p = 1$  so that  $\frac{n-p-1}{c(n-1)} + \frac{p}{n-1} = \frac{n-2-1}{c(n-1)} + \frac{1}{n-1}$  and as  $c \in \left\{ \frac{n(n-2)}{k(n-1)-n} \mid k \in \mathbb{N}, 2 \leq k \leq \lfloor n/2 \rfloor; k/2 = \lfloor k/2 \rfloor \right\}$  we have

$$\frac{n-2-1}{c(n-1)} + \frac{1}{n-1} = \frac{k}{n}$$

where  $k$  is an even integer larger than 2 and smaller than  $\lfloor n/2 \rfloor$ . When  $n$  gets large, the range of possible values for the level of the test that may be exactly reached increases. Provide  $n$  is large enough we can find a distribution under the null such that the size of our exact Fisher procedure is as close as desired to any  $\alpha \in [0, 1/2]$ .

In the above example, for large values of  $n$  one “disturbance” term is close to 1 and all the others are close to zero. This may viewed (up to a re-normalization) as the realization of an arbitrarily large “outlier” for one observation in an iid setting. As the –unique– explanatory variable is a dummy that splits the sample in three parts, the test statistic can only take two values corresponding to the case where the “outlier” belongs to the first or the second part of the sample. Finally, it should be notice that in the above example error terms take values in a set with known boundaries. This case then fulfills the assumptions by Gossner and Schlag [2013].

## 2.5 Multivariate moments of exchangeable self-centered self-normalized processes.

If  $\eta$  is a random vector such that  $\sum_{i=1}^n \eta_i^2 = 1$  every component of this vector is bounded. Hence all moments of  $\eta$  exist. Proposition 1 has been derived using the two first moments only. We now investigate improvement that may follow from using higher order versions of the Markov bound.

Several caveats must be stressed. First, as Proposition 2 makes clear, such improvements cannot be “universally strict”. Second, every exchangeable process  $\eta$  such that  $\sum_{i=1}^n \eta_i = 0, \sum_{i=1}^n \eta_i^2 = 1$  share the same two first moments, but this is no longer the case for higher order moments, as shown in the following example.

Consider  $\eta$  distributed uniformly among the permutations of the vector  $(0, 1/\sqrt{2}, -1/\sqrt{2})$ . We have  $E_\eta[\eta_1^3] = 0$ . Now consider  $\eta$  distributed uniformly among the permutations of the vector  $(\sqrt{2/3}, -1/\sqrt{6}, -1/\sqrt{6})$ . We have  $E_\eta[\eta_1^3] = 1/(3\sqrt{6})$ .

This implies that higher order of bound lead to critical points that depend on some unknown moments. A genuine control of the type-I error requires a maximization over all possible values of this moments. We will now discuss theoretical and practical questions related to this problem.

First notice the Fisher statistic is a quadratic function of  $\eta$  so that it is invariant by considering  $-\eta$  instead of  $\eta$ . This means that we may restrict ourselves to even multivariate moments.<sup>15</sup> Now by symmetry all odd moments of symmetrical processes are zero.

<sup>15</sup>More formally, let  $\mathcal{H}$  be the set of possible distribution for exchangeable, self-centered self-normalized random vector  $\eta$ . If we associate to each process  $\eta$  whose distribution belongs to  $\mathcal{H}$  a symmetric process  $\eta^*$  obtained by  $\mathbf{P}(\eta^* = \eta) = \mathbf{P}(\eta^* = -\eta) = 1/2$ . Conditionally on  $X$ , the Fisher statistic has the same distribution for these two processes. Hence the supremum over all distributions is the same as the supremum over all symmetrical distributions.

Second, thanks to our exchangeability condition, the number of different multivariate order- $k$  moments is equal to the  $p(k)$  the number of integer partitions of  $k$ . In particular, if we are about to use a ' $k$ -order version' of the Markov bound (that is to say using the inequality involving a moment of order  $k$ ), the number of relevant moments does not increase with the sample size.

The function  $k \rightarrow p(k)$  is well known in combinatorics. For instance  $p(4) = 5$ ,  $p(6) = 11$  and  $p(8) = 22$ . In our setting, the additional restrictions  $\sum_i^n \eta_i = 0$  and  $\sum_i^n \eta_i^2 = 1$  provide several relationships that may be used to reduce further the number of nuisance parameters. As  $\eta$  is a size- $n$  exchangeable process, all of the  $\frac{n!}{(n-k)!}$   $k$ -dimensional conditional distributions of  $\eta$  given  $X$  are identical, for  $k = 1, \dots, (n-1)$ . This implies that the fourth-order moments fulfills for any  $n > 4$

$$\begin{aligned} 0 &= E_\eta [\eta_1^3 \sum_{i=1}^n \eta_i] = E_\eta[\eta_1^4] + (n-1)E_\eta[\eta_1^3 \eta_2] \\ 0 &= E_\eta [\eta_1^2 \eta_2 \sum_{i=1}^n \eta_i] = E_\eta[\eta_1^3 \eta_2] + E_\eta[\eta_1^2 \eta_2^2] + (n-2)E_\eta[\eta_1^2 \eta_2 \eta_3] \\ 0 &= E_\eta [\eta_1 \eta_2 \eta_3 \sum_{i=1}^n \eta_i] = 3E_\eta[\eta_1^2 \eta_2 \eta_3] + (n-3)E_\eta[\eta_1 \eta_2 \eta_3 \eta_4] \\ n^{-1} &= E_\eta [\eta_1^2 \sum_{i=1}^n \eta_i^2] = E_\eta[\eta_1^4] + (n-1)E_\eta[\eta_1^2 \eta_2^2] \\ -(n(n-1))^{-1} &= E_\eta [\eta_1 \eta_2 \sum_{i=1}^n \eta_i^2] = 2E_\eta[\eta_1^3 \eta_2] + (n-2)(n-3)E_\eta[\eta_1^2 \eta_2 \eta_3] \end{aligned}$$

This is a linear system with 5 equations and 5 variables. All order-4 moments may be explicitly computed from  $E_\eta[\eta_1^4]$ . More precisely, we have

$$\begin{aligned} E_\eta[\eta_1^3 \eta_2] &= -\frac{E_\eta[\eta_1^4]}{n-1} \\ E_\eta[\eta_1^2 \eta_2^2] &= \frac{1-nE_\eta[\eta_1^4]}{n(n-1)} \\ E_\eta[\eta_1^2 \eta_2 \eta_3] &= \frac{2nE_\eta[\eta_1^4]-1}{n(n-1)(n-2)} \\ E_\eta[\eta_1 \eta_2 \eta_3 \eta_4] &= \frac{3-6nE_\eta[\eta_1^4]}{n(n-1)(n-2)(n-3)} \end{aligned}$$

Proceeding accordingly for moments of every order we see that the relevant  $p(k)$  multivariate moments of order  $k$  are solutions of a linear system with  $p(k-1) + p(k-2)$  equations.

The minimal number of moments involved in the computation of the expectation of all moments is given by the next proposition<sup>16</sup>

**Proposition 3** *Consider for all size- $n$  exchangeable process  $\eta$  such that  $\sum_{i=1}^n \eta_i = 0$  and  $\sum_{i=1}^n \eta_i^2 = 1$ . The vector space spanned by the set  $\{E_\eta[Q(\eta)] | Q \in \mathcal{Q}_n(d)\}$  where  $\mathcal{Q}_n(d)$  is the set of  $n$  variables homogeneous polynomial of degree  $d$ , is in the vector space spanned by the set*

$$\left\{ E_\eta \left[ \prod_{i=3}^n \eta_i^{a_i} \right] \mid (a_3, \dots, a_n) \in \mathbb{N}^n; a_3 \geq a_4 \geq \dots \geq a_n; \sum_{i=3}^n i \times a_i = d \right\} \cup \{1\}.$$

*Proof :*

Let  $S_n$  the symmetric group on  $\{1, \dots, n\}$ ,  $\mathcal{Q}_n$  the set of polynomial from  $\mathbb{R}^n$  to  $\mathbb{R}$  and  $\mathcal{Q}_n(d)$  the subset of  $\mathcal{Q}_n$  that contains homogeneous of degree- $d$  polynomials.

<sup>16</sup>We thank Jean-Yves Welschinger for providing us this result.

For any polynomial  $Q \in \mathcal{Q}_n$  define

$$\text{Sym}(Q)(u_1, \dots, u_n) = \frac{1}{n!} \sum_{\sigma \in S_n} Q(u_{\sigma(1)}, \dots, u_{\sigma(n)})$$

The polynomial  $\text{Sym}(Q)$  is symmetrical. Also  $\text{Sym}^2 = \text{Sym}$  hence  $\text{Sym}$  is a projector on the set of *symmetric* polynomials in  $\mathcal{Q}_n$ .

The set  $\mathcal{Q}_n$  may then be decomposed into direct sum  $\text{Im}(\text{Sym}) \oplus \text{Ker}(\text{Sym})$ .

If  $\eta$  is a size- $n$  exchangeable process then for  $Q \in \mathcal{Q}_n$  we have  $E(Q(\eta)) = E(Q(\eta_\sigma))$  for any  $\sigma \in S_n$ . Thus  $E_\eta[Q(\eta)] = E_\eta[\text{Sym}(Q)(\eta)]$ . It follows

$$Q \in \text{Ker}(\text{Sym}) \Rightarrow E_\eta[Q(\eta)] = 0.$$

As a consequence for any  $Q \in \mathcal{Q}_n$  we may write

$$E_\eta[Q(\eta)] = E_\eta[\text{Sym}(Q)(\eta)]$$

Now the Newton identities (see, for instance, Kalman [2000]) state that if  $Q \in \mathcal{Q}_n(d)$ , then  $\text{Sym}(Q)(\eta)$  may be written as a linear combination of

$$\left( \sum_{i=1}^n \eta_i^k \right)_{k \in \{0, 1, \dots, d\}}$$

This implies that when  $\eta$  is exchangeable we may write

$$E_\eta[Q(\eta)] = E_\eta \left[ R \left( \sum_{i=1}^n \eta_i, \sum_{i=1}^n \eta_i^2, \sum_{i=1}^n \eta_i^3, \sum_{i=1}^n \eta_i^4, \dots, \sum_{i=1}^n \eta_i^d \right) \right]$$

for some  $R \in \mathcal{Q}_d$

The conditions  $0 = \sum_{i=1}^n \eta_i$  and  $1 = \sum_{i=1}^n (\eta_i)^2$  implies further simplification. Namely, the existence of a new polynomial  $S$  such that

$$E_\eta[Q(\eta)] = E_\eta \left[ S \left( \sum_{i=1}^n \eta_i^3, \sum_{i=1}^n \eta_i^4, \dots, \sum_{i=1}^n \eta_i^d \right) \right]$$

Now consider a monomial function

$$M_{a_3, \dots, a_n}(t_3, \dots, t_n) \rightarrow \prod_{j=3}^n t_j^{a_j}$$

of degree  $\sum_{j=3}^n a_j$ . The degree of  $M_{a_3, \dots, a_n}(\sum_{i=1}^n \eta_i^3, \dots, \sum_{i=1}^n \eta_i^n)$  as a polynomial function of  $\eta$  is

$$3a_3 + 4a_4 + \dots + na_n$$

As a consequence, if we consider  $E_\eta[Q(\eta)]$  for some homogeneous polynomial of degree  $d$ , it may be written as a linear combination of 1 and  $E_\eta[\prod_{i=3}^n \eta_i^{a_i}]$  under the constraints

$$\begin{aligned} \sum_{i=3}^n i \times a_i &= d \\ a_3 &\geq a_4 \geq \dots \geq a_n \end{aligned}$$

where the last sequence of inequalities use again exchangeability.

*Q.E.D*

The previous proposition allows us to present the moments needed to compute the expectation of  $E_\eta[Q(\eta)]$  for any homogeneous, degree- $d$  polynomial for the first 10 moments

degree	moments
3	$E_\eta[\eta_1^3]$
4	$E_\eta[\eta_1^4]$
5	$E_\eta[\eta_1^5]$
6	$E_\eta[\eta_1^6]; E_\eta[\eta_1^3\eta_2^3]$
7	$E_\eta[\eta_1^7]; E_\eta[\eta_1^4\eta_2^3]$
8	$E_\eta[\eta_1^8]; E_\eta[\eta_1^5\eta_2^3]; E_\eta[\eta_1^4\eta_2^4]$
9	$E_\eta[\eta_1^9]; E_\eta[\eta_1^4\eta_2^5]; E_\eta[\eta_1^3\eta_2^3\eta_3^3]$
10	$E_\eta[\eta_1^{10}]; E_\eta[\eta_1^7\eta_2^3]; E_\eta[\eta_1^6\eta_2^4]; E_\eta[\eta_1^5\eta_2^5]; E[\eta_1^4\eta_2^3\eta_3^3]$

As we see, the moments to consider are related to the partitions of  $d$  that do not contain 1 nor 2. The sequence corresponding to the number of such partitions as a function of  $d$  is labeled A008483 in the On-line Encyclopedia of Integer Sequences [\[R\]](#).

The following proposition provide a recursive way to compute the number of such partitions.

**Proposition 4** *Let  $p_{>2}(k, n)$  be the number of partitions of an integer  $n$  into  $k$  parts the smaller of which is strictly larger than 2. We have for all couple of integers  $k, n$*

$$\begin{aligned}
p_{>2}(1, n) &= 1 \\
p_{>2}(k, n) &= 0 \quad \text{if} \quad k > \lfloor n/3 \rfloor \\
p_{>2}(\lfloor n/3 \rfloor, n) &= 1 \\
p_{>2}(k, n) &= p_{>2}(k, n - k) + p_{>2}(k - 1, n - 3)
\end{aligned}$$

The first and second equality are trivial. For the third one, we have  $n = 3\lfloor n/3 \rfloor + i$  with  $i \in \{0, 1, 2\}$  and there is only one way to divide  $n$  into  $\lfloor n/3 \rfloor$  parts all of which are strictly larger than 2, namely  $(3, 3 \dots 3, 3 + i)$ .

For the last one consider a partition of  $n$  into  $k$  parts all of which are strictly larger than 2. If its smallest element equals 3 then it may be viewed as a partition of  $n - 3$  into  $k - 1$  parts all of which are strictly larger than 2 to which '3' has been concatenated. Otherwise, the smallest element of this partition is at least 4. In such a case, this partition may be formed by adding 1 to each element of a partition of  $n - k$  into  $k$  parts all of which are strictly larger than 2. Q.E.D

The number of moments is rapidly increasing. For instance consider increasing the degree by one: if  $d + 1 = 20$  they are "extra" 49 moments to consider, whereas if  $d + 1 = 30$  leads to 331 extra moments <sup>17</sup>.

It may also be shown that the number of partitions that do not contain 1 nor 2 as exactly the same asymptotic as  $p(n)$  namely  $\exp\left(\pi\sqrt{\frac{2n}{3}}\right)$  up to some

<sup>17</sup>This requires that  $n$  is larger than  $d + 1$ . Clearly, if  $n = 3$  the number of moments is much weaker.



constant term. This follows from the fact that  $\mathbb{N} \setminus \{1, 2\}$  as a natural density equal to 1 (for details, see Nathanson [2000] p. 495).<sup>18</sup> This suggests that the gain in efficiency we can hope by considering higher moment versions of the Markov inequality rapidly vanishes as the degree increases. Moreover, as showed in the next subsection, using very high order moments implies an increasing computational burden. In this paper, we shall restrict ourselves to moments up to order 8.<sup>19</sup>

Several constraints for these moments may also be derived. Using the Cauchy-Schwarz inequality we get

$$\frac{1}{n^2} = (E_\eta[\eta_1^2])^2 \leq E_\eta[\eta_1^4] \leq E_\eta[\eta_1^2] = \frac{1}{n}$$

where the inequality on the right hand side follows from  $\eta_1^2 < 1$ .

The above lower bound is tight. Assume  $n$  is even, consider equiprobable permutations of the vector  $(n^{-1/2}, -n^{-1/2}, n^{-1/2}, -n^{-1/2}, \dots, n^{-1/2})$  we have  $E_\eta[\eta_1^4] = n^{-2}$ . In large samples, we may be arbitrarily close to the upper bound if consider the process used in the proof of Proposition 2 above.

Also, as odd moments may all be considered as zero, we have

$$E_\eta[\eta_1^3 \eta_2^3] = \text{Cov}_\eta[\eta_1^3; \eta_2^3] < 0$$

since  $E_\eta[\eta_1 \eta_2] = \text{Cov}_\eta[\eta_1; \eta_2] < 0$  and  $x \rightarrow x^3$  is an increasing function.

As for  $E_\eta[\eta_1^6]$  we have

$$(E_\eta[\eta_1^4])^2 \leq E_\eta[\eta_1^6] \leq E_\eta[\eta_1^4]$$

where the LHS follows from  $(E_\eta[\eta_1^4])^2 \leq E_\eta[\eta_1^8] \leq E_\eta[\eta_1^6]$ .

All these inequalities may be used to derive exact procedures based on higher order versions of the Markov inequality.

## 2.6 Higher-order Markov inequalities : practical issues

We now address computational issues related to the derivation of such bounds. First consider using the boundary

$$\mathbf{P}_\eta(\eta' B \eta > \kappa | X) \leq \frac{E_\eta[(\eta' B \eta)^2 | X]}{\kappa^2}.$$

We know that  $E_\eta[(\eta' B \eta)^2 | X]$  depends only on order-4 multivariate moments of  $\eta$  so that there exist two constants depending on  $B$  only such that  $E_\eta[(\eta' B \eta)^2 | X] = a(B) + b_4(B)E_\eta[\eta_1^4]$ . The notation  $a(B)$  and  $b_4(B)$  stress that these constants depends on the realization of the exogenous variables  $X$  and the constraints to be tested.

For a given matrix  $B$ , the quantities  $a(B)$  and  $b_4(B)$ , may be derived by computing  $E_\eta[(\eta' B \eta)^2 | X]$  in two specific cases for the process  $\eta$ . From a computational standpoint, a quick way to do it is to use equiprobable permutations

<sup>18</sup>Another direct link between  $p(k, n)$  and  $p_{>2}(k, n)$  is as follows. Consider a partition  $a_1, \dots, a_k$  of the integer  $n > 2k$  that does not contain neither 1 nor 2. Then  $a_1 - 2, \dots, a_k - 2$  is a partition of  $n - 2k$  with  $k$  elements. Similarly, if  $b_1, \dots, b_k$  is a partition of  $k$  elements of the integer  $n$  then  $2 + b_1, \dots, 2 + b_k$  is a partition of the integer  $n + 2k$  with  $k$  elements. Hence we have  $p(n + 2k, k) = p_{>2}(k, n)$ . We thank Philippe Barbe for this remark.

<sup>19</sup>In the current section we make use of order-6 moments only. Order-8 moments will be necessary in Section 3.

of the vector  $(x, y, y, \dots, y)$  with  $x = \sqrt{(n-1)/n}$ ,  $y = -(n(n-1))^{-1/2}$  and equiprobable permutations of a vector  $(x, y, z, z, \dots, z)$  under the constraints  $x + y + (n-2)z = 0$ ,  $x^2 + y^2 + (n-2)z^2 = 1$ . The computational cost is of order  $n^2$ .<sup>20</sup>

It must be stressed that although our approach bears resemblance with permutation testing, there are important differences. First, we do not use permutation of the *original* data set but of a fictitious one in which "disturbances" have a given distribution. In particular, this new data set does not depend on  $Y$ . Second the permutation is used solely for computational simplicity. If the matrix  $B$  is such that some clever choice of the distribution of  $\eta$  allows for a direct computation of  $a(B)$  and  $b_4(B)$ , one should use it. Finally, whereas permutation tests may typically be carried out using a random sampling of permutations, our exact approach rests on the examination of *all* possible permutations. The above choices of distributions allows us to avoid the intractable computations of  $n!$  versions of the same statistic.

When we have derived  $a(B)$  and  $b_4(B)$ , we must compute the bound as a solution of the following problem

$$\begin{aligned} \max \quad & a(B) + b_4(B)E_\eta[\eta_1^4] \\ \text{wrt} \quad & E_\eta[\eta_1^4] \\ \text{s.t.} \quad & n^{-2} \leq E_\eta[\eta_1^4] \leq n^{-1} \end{aligned}$$

If we want to use a bound depending on  $E_\eta[(\eta' B \eta)^3 | X]$  we will need at least 4 different distributions since proposition 3 above implies

$$E_\eta[(\eta' B \eta)^3 | X] = a'(B) + b'_4(B)E_\eta[\eta_1^4] + b_6(B)E_\eta[\eta_1^4] + b_{33}(B)E_\eta[\eta_1^6]$$

Unfortunately, using equiprobable permutations of a vector  $(x, y, z, z, \dots, z)$  under the constraints  $x + y + (n-2)z = 0$ ,  $x^2 + y^2 + (n-2)z^2 = 1$  is no longer sufficient, as we have in this case

$$\begin{aligned} E_\eta[\eta_1^4] &= \frac{1}{2n} \\ E_\eta[\eta_1^6] &= \frac{12x^6 - 12x^4 + 3x^2 + 1}{4n} \\ E_\eta[\eta_1^3 \eta_2^3] &= \frac{24x^6 - 24x^4 + 6x^2 - 1}{4n(n-1)} \end{aligned}$$

Hence in particular  $2E_\eta[\eta_1^6] = (n-1)E_\eta[\eta_1^3 \eta_2^3]$  and such distributions cannot identify  $b'_4(B)$ ,  $b_6(B)$ .

We then need to consider vectors of the form  $(x, y, z, w, w, \dots, w)$  under the constraints  $x + y + z + (n-3)w = 0$ ,  $x^2 + y^2 + z^2 + (n-3)w^2 = 1$ . The cost of this computation raises to  $n^3$ .

A valid bound then derives from the following problem

---

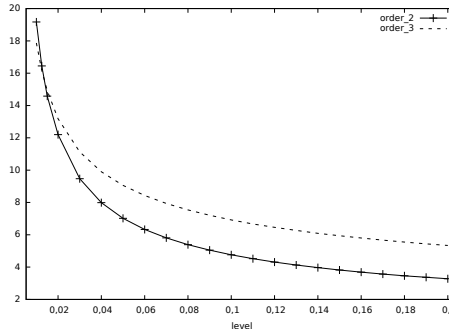
<sup>20</sup>Take care of the fact that the computation of  $a(B)$  and  $b_4(B)$  does not require the process to be symmetrical. Symmetry is used in the optimization step, see below.

$$\begin{aligned}
& \max && a'(B) + b'_4(B)E_\eta[\eta_1^4] + b_6(B)E_\eta[\eta_1^6] + b_{33}(B)E_\eta[\eta_1^3\eta_2^3] \\
& \text{wrt} && \\
& E_\eta[\eta_1^4], E_\eta[\eta_1^6], E_\eta[\eta_1^3\eta_2^3] \\
& \text{s.t.} && n^{-2} \leq E_\eta[\eta_1^4] \leq n^{-1} \\
& && E_\eta[\eta_1^6] \leq E_\eta[\eta_1^4] \leq 0 \\
& && -E_\eta[\eta_1^6] \leq E_\eta[\eta_1^3\eta_2^3] \leq 0
\end{aligned}$$

This problem is a simplex in dimension 3 that is solvable by direct inspection of the edges. Contrarily to the previous maximization problem, it is not clear that all edges corresponds to a feasible set of order-6 moments. Hence, the second maximization problem may not lead to a better bound.

The following chart gives a comparison of the critical points derived from order-2 and order-3 moments inequalities for a sample size  $n = 50$  with a single linear trend explanatory variable (this case is considered for the sake of illustration).

Figure 1: Comparison of critical points derived from order-3 and order-2 inequalities (linear trend case:  $E[Y_t] = a + bt$ )



We see that the second maximization provides a better bound for very small levels only. Higher order Markov bounds may nevertheless provide a much better test. For instance, at level 5%, the critical point associated with the decision rule derived from the Markov bound presented in Section 3.1 above is 33.10, whereas the best of the two higher order tests is 7.02.

For such a sample size, the Student critical point for a 5% size in the Gaussian case is approximately 2.31. Yet the size of this test in our model is bounded from below by 38.59% (using the distribution presented in the proof of Proposition 2).

### 3 Exact testing in an ARX(1) model with exchangeable error terms

In this section we consider the very same set-up as Dufour and Kiviet [1998] (hereafter DK). The purpose of this section is to remove –their– assumption B page 82 according to which the distribution of error is known up to some scalar and replace it by the weaker assumption that the error terms are exchangeable conditionally on the exogenous explanatory variables.

#### 3.1 An AR(1) model with exogenous variables and exchangeable error terms

We consider a family  $\mathcal{D}$  of probability distributions for  $(Y, X)$ , where  $Y, X$  are matrices of real random variables of size  $n \times 1$  and  $n \times p$  respectively (with  $p < n$ ). In this section we assume the following condition holds

**Assumption 2** *For any  $\mathbf{P} \in \mathcal{D}$  there exist a single size- $p$  vector denoted  $\beta(\mathbf{P})$  and a single real  $\rho(\mathbf{P})$  such that the distribution of  $Y - \rho(\mathbf{P})Y_{-1} - X\beta(\mathbf{P})$  conditional on  $X, y_0$  is exchangeable and non degenerated.*

We consider the null hypothesis  $H_0 : \rho(\bar{\mathbf{P}}) = \rho_0, R\beta(\bar{\mathbf{P}}) = \gamma_0$ . As before, we write  $\mathbf{P} \in \mathcal{D}_0$  whenever  $\rho(\mathbf{P}) = \rho_0$  and  $R\beta(\mathbf{P}) = \gamma_0$ .

First as in DK, let  $\bar{R}$  be a  $(p - q) \times p$  matrix such that  $Q = (R'|\bar{R}')'$  is an invertible  $p \times p$  matrix. Also denote  $Z = (Z_1|Z_2) = XQ^{-1}$  where the partition of  $Z$  is conformable to that of  $Q$ . In the case  $q = 0$  we simply consider  $Z_2 = X$  and  $Z_1$  is void. Similarly, if  $q = p$  then  $Z_2$  is void.

The above assumption may be rewritten as:

**Assumption 3** *For any  $\mathbf{P} \in \mathcal{D}$  there exist a single size- $q$  vector denoted  $\gamma_1(\mathbf{P})$ , a single size  $p - q$  vector  $\gamma_2(\mathbf{P})$  and a single real  $\rho(\mathbf{P})$  such that the distribution of  $Y - \rho(\mathbf{P})Y_{-1} - Z_1\gamma_1(\mathbf{P}) - Z_2\gamma_2(\mathbf{P})$  conditional on  $Z, y_0$  is exchangeable and non degenerated.*

The above change in variables and the induced new parameterization leaves us with a new writing of the null hypothesis  $H_0 : \rho(\mathbf{P}) = \rho_0, \gamma_1(\mathbf{P}) = \text{gamma}_0$ . DK use the change of variables to highlight the fact that the (pseudo)-Likelihood Ratio test associated with  $H_0$  is in general not pivotal ( $\gamma_2(\mathbf{P})$  is a nuisance parameter). They propose several other statistics that may be viewed as (pseudo) LR tests in an “enlarged” model. This amounts to include extra explanatory variables in the original equation. The model  $\mathcal{D}$  appears as a special case of this enlarged model in which the coefficients associated with extra variables are set to zero. DK then consider (pseudo) LR statistic related to the hypothesis  $H_0 : \rho(\bar{\mathbf{P}}) = \rho_0, R\beta(\bar{\mathbf{P}}) = \gamma_0$  in the enlarged model. Under the null this statistic depends only on the distribution of the process  $Y - \rho(\mathbf{P})Y_{-1} - X\beta(\mathbf{P})$  (see DK for details).

We introduce the following notations. For any  $A$  full column rank matrix define  $M_A = I - A(A'A)^{-1}A'$ . Also define

$$C(\rho_0) = \begin{pmatrix} 0 & . & . & . & . & . & 0 \\ 1 & 0 & . & . & . & . & 0 \\ \rho_0 & 1 & 0 & . & . & . & 0 \\ \rho_0^2 & \rho_0 & 1 & . & . & . & 0 \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ \rho_0^{n-2} & \rho_0^{n-3} & . & . & \rho_0 & 1 & 0 \end{pmatrix}, \quad \iota(\rho_0) = \begin{pmatrix} 1 \\ \rho_0 \\ \rho_0^2 \\ . \\ . \\ . \\ \rho_0^{n-1} \end{pmatrix}$$

One of the test statistic considered by DK is

$$\frac{(Y - \rho_0 Y_{-1} - Z_1 \gamma_0)' M_N (Y - \rho_0 Y_{-1} - Z_1 \gamma_0)}{Y' M_D Y}$$

where

$$\begin{aligned} N &= (Z_2 | \iota(\rho_0) | C(\rho_0) Z_2 | C(\rho_0) Z_1 \gamma_0) \\ D &= (Z | \iota(\rho_0) | C(\rho_0) Z_2 | C(\rho_0) Z_1 \gamma_0 | Y_{-1}) \end{aligned}$$

Now, for all  $\mathbf{P} \in \mathcal{D}$  define  $\epsilon(\mathbf{P}) = Y - \rho(\mathbf{P}) Y_{-1} - X \beta(\mathbf{P})$ . DK have shown that (see the proof of their property 1)

$$\mathbf{P} \in \mathcal{D}_0 \Rightarrow \begin{cases} (Y - \rho_0 Y_{-1} - Z_1 \gamma_0)' M_N (Y - \rho_0 Y_{-1} - Z_1 \gamma_0) = \epsilon'(\mathbf{P}) M_N \epsilon(\mathbf{P}) \\ Y' M_D Y = \epsilon'(\mathbf{P}) M_E \epsilon(\mathbf{P}) - \frac{(\epsilon'(\mathbf{P}) M_E C(\rho_0) \epsilon(\mathbf{P}))^2}{\epsilon'(\mathbf{P}) C'(\rho_0) M_E C(\rho_0) \epsilon(\mathbf{P})} \end{cases}$$

with  $E = (Z | \iota(\rho_0) | C(\rho_0) Z_2 | C(\rho_0) Z_1 \gamma_0)$ .

As  $\iota_n$  belongs to the vector space spanned by the columns of matrix  $N$  and  $E$  and  $C(\rho_0) \iota_n$  belongs to the vector space spanned by the columns of matrix  $E$ , we may write for all  $\mathbf{P} \in \mathcal{D}_0$

$$\begin{aligned} (Y - \rho_0 Y_{-1} - Z_1 \gamma_0)' M_N (Y - \rho_0 Y_{-1} - Z_1 \gamma_0) &= u'(\mathbf{P}) M_N u(\mathbf{P}) \\ Y' M_D Y &= u'(\mathbf{P}) M_E u(\mathbf{P}) - \frac{(u'(\mathbf{P}) M_E C(\rho_0) u(\mathbf{P}))^2}{u'(\mathbf{P}) C'(\rho_0) M_E C(\rho_0) u(\mathbf{P})} \end{aligned}$$

where  $u(\mathbf{P}) = M_{\iota_n} \epsilon(\mathbf{P})$  and the test statistic is found to be distributed under the null as

$$\frac{\eta' M_N \eta}{\eta' M_E \eta - \frac{(\eta' M_E C(\rho_0) \eta)^2}{\eta' C'(\rho_0) M_E C(\rho_0) \eta}}$$

where  $\eta$  is again an exchangeable vector such that  $\sum_{i=1}^n \eta_i = 0$  and  $\sum_{i=1}^n \eta_i^2 = 1$ .

### 3.2 Bound deriving from the Cantelli inequality

Contrarily to DK, the above statistic cannot be simulated since the distribution of the error terms is unknown. We thus seek for a bound to the (conditional) probability of the following event

$$\frac{\eta' M_N \eta}{\eta' M_E \eta - \frac{(\eta' M_E C(\rho_0) \eta)^2}{\eta' C'(\rho_0) M_E C(\rho_0) \eta}} \geq k$$

or (for positive values of  $k$ )

$$\eta' \left( \frac{1}{k} M_N - M_E \right) \eta \eta' C'(\rho_0) M_E C(\rho_0) \eta + (\eta' M_E C(\rho_0) \eta)^2 \geq 0 \quad (4)$$

when  $\eta$  is an exchangeable vector such that  $\sum_{i=1}^n \eta_i = 0$  and  $\sum_{i=1}^n \eta_i^2 = 1$ .

Let  $T(k, \eta) = \eta' \left( \frac{1}{k} M_N - M_E \right) \eta \eta' C'(\rho_0) M_E C(\rho_0) \eta + (\eta' M_E C(\rho_0) \eta)^2$ . It is a square integrable variable as a fourth-degree polynomial function of  $\eta$ .

Using the particular form  $T(k, \eta)$  as a function of  $\eta$ , the following Corollary of Proposition 3 provides important results to control the probability of the event (4).

**Corollary 1**

i) There exist a vector  $v = (v_1, \dots, v_4) \in \mathbb{R}^4$  depending only on  $X, y_0$  and on  $\rho_0, R, \gamma_0$  such that  $\sum_{i=1}^n \eta_i = 1 - \sum_{i=1}^n \eta_i^2 = 0$  and  $\eta$  exchangeable imply

$$E_\eta[T(k, \eta)|X, y_0] = v_1 + v_2/k + E_\eta[\eta_1^4](v_3 + v_4/k)$$

ii) There exists a vector  $w = (w_1, \dots, w_{21}) \in \mathbb{R}^{21}$  depending only on the realization of explanatory variables and on  $\rho_0, R, \gamma_0$  such that  $\sum_{i=1}^n \eta_i = 1 - \sum_{i=1}^n \eta_i^2 = 0$  and  $\eta$  exchangeable imply

$$\left\{ \begin{array}{l} E_\eta[T^2(k, \eta)|X] = w_1 + (w_2 + w_3/k)/k \\ \quad + E_\eta[\eta_1^4](w_4 + (w_5 + w_6/k)/k) \\ \quad + E_\eta[\eta_1^6](w_7 + (w_8 + w_9/k)/k) \\ \quad + E_\eta[\eta_1^8](w_{10} + (w_{11} + w_{12}/k)/k) \\ \quad + E_\eta[\eta_1^3 \eta_2^3](w_{13} + (w_{14} + w_{15}/k)/k) \\ \quad + E_\eta[\eta_1^4 \eta_2^4](w_{16} + (w_{17} + w_{18}/k)/k) \\ \quad + E_\eta[\eta_1^5 \eta_2^3](w_{19} + (w_{20} + w_{21}/k)/k) \end{array} \right.$$

We can now state the following result

**Proposition 5** Let  $v, w$  be the vectors described in corollary 1 above. Assume  $k$  is such that  $v_1 + v_2/k + \max\{(v_3 + v_4/k)n^{-2}; (v_3 + v_4/k)n^{-1}\} \leq 0$ . then

$$\sup_{\mathbf{P} \in \mathcal{D}_0} \mathbf{P} \left( \frac{(Y - \rho_0 Y_{-1} - Z_1 \gamma_0)' M_N (Y - \rho_0 Y_{-1} - Z_1 \gamma_0)}{Y' M_D Y} > k | X, y_0 \right)$$

is less or equal to

$$\left\{ \begin{array}{l} \text{argmax } 1 - \frac{U}{L} \\ \text{w.r.t} \\ E_\eta[\eta_1^4], E_\eta[\eta_1^6], E_\eta[\eta_1^8], E_\eta[\eta_1^3\eta_2^3], E_\eta[\eta_1^4\eta_2^4], E_\eta[\eta_1^5\eta_2^3] \\ \text{s.t.} \\ U = v_1 + v_2/k + E_\eta[\eta_1^4](v_3 + v_4/k) \\ L = w_1 + (w_2 + w_3/k)/k \\ \quad + E_\eta[\eta_1^4](w_4 + (w_5 + w_6/k)/k) \\ \quad + E_\eta[\eta_1^6](w_7 + (w_8 + w_9/k)/k) \\ \quad + E_\eta[\eta_1^8](w_{10} + (w_{11} + w_{12}/k)/k) \\ \quad + E_\eta[\eta_1^3\eta_2^3](w_{13} + (w_{14} + w_{15}/k)/k) \\ \quad + E_\eta[\eta_1^4\eta_2^4](w_{16} + (w_{17} + w_{18}/k)/k) \\ \quad + E_\eta[\eta_1^5\eta_2^3](w_{19} + (w_{20} + w_{21}/k)/k) \\ n^{-1} \geq E_\eta[\eta_1^4] \geq n^{-2} \\ E_\eta[\eta_1^4] \geq E_\eta[\eta_1^6] \geq E_\eta[\eta_1^8] \geq (E_\eta[\eta_1^4])^2 \\ 0 \geq E_\eta[\eta_1^3\eta_2^3] \geq -E[\eta_1^6] \\ 0 \geq E_\eta[\eta_1^5\eta_2^3] \geq -E[\eta_1^8] \\ E_\eta[\eta_1^4\eta_2^4] \leq E_\eta[\eta_1^8] \end{array} \right. \quad (5)$$

*Proof:* For any  $\mathbf{P} \in \mathcal{D}_0$ , the conditional probability  $\mathbf{P}_\eta(T(k, \eta) > 0 | X, y_0)$  equals

$$\mathbf{P} \left( \frac{(Y - \rho_0 Y_{-1} - Z_1 \gamma_0)' M_N (Y - \rho_0 Y_{-1} - Z_1 \gamma_0)}{Y' M_D Y} > k | X, y_0 \right) =$$

As  $E_\eta[T(k, \eta) | X, y_0]$  is a linear function of  $E_\eta[\eta_1^4]$ , and  $n^{-1} \geq E_\eta[\eta_1^4] \geq n^{-2}$  the condition  $v_1 + v_2/k + \max\{(v_3 + v_4/k)n^{-2}; (v_3 + v_4/k)n^{-1}\} \leq 0$  implies  $E_\eta[T(k, \eta) | X, y_0] < 0$ . The Cantelli inequality then implies

$$\mathbf{P}_\eta(T(k, \eta) > 0 | X, y_0) \leq \frac{\text{Var}_\eta[T(k, \eta) | X, y_0]}{E_\eta[T(k, \eta)^2 | X, y_0]} = 1 - \frac{E_\eta[T(k, \eta) | X, y_0]^2}{E_\eta[T(k, \eta)^2 | X, y_0]}$$

As  $\frac{E_\eta[T(k, \eta) | X, y_0]^2}{E_\eta[T(k, \eta)^2 | X, y_0]}$  depends on the unknown moments

$$E_\eta[\eta_1^4], E_\eta[\eta_1^6], E_\eta[\eta_1^8], E_\eta[\eta_1^3\eta_2^3], E_\eta[\eta_1^4\eta_2^4], E_\eta[\eta_1^5\eta_2^3]$$

a feasible boundary requires optimization of the boundary wrt these quantities.

As  $T(k, \eta)$  is unchanged if we change  $\eta$  to  $-\eta$ , we may consider the case of symmetrical self-centered self-normalized exchangeable process  $\eta$ . Finally, proceeding as in section 2.5, it is not difficult to show that for such a process the quantities  $E_\eta[\eta_1^4], E_\eta[\eta_1^6], E_\eta[\eta_1^8], E_\eta[\eta_1^3\eta_2^3], E_\eta[\eta_1^4\eta_2^4], E_\eta[\eta_1^5\eta_2^3]$  fulfill the constraints described in program (5) <sup>21</sup>. Q.E.D

<sup>21</sup>Note other constraints may be added. Using computations made in Section 2.5, together with the Cauchy-Schwarz inequality, we get

### 3.3 Implementation issues

Using the above result in practice raise difficult computational problems. First, we have to compute the vectors  $v$  and  $w$  for the problem under study.

Consider first  $v$ . As in section 2.6 we may compute exactly  $E_\eta[T(k, \eta)|X, y_0]$  for  $k = 1$  and  $k = 2$  (say) using equiprobable permutations of the vectors  $(\sqrt{(n-1)/n}, -(n(n-1))^{-1/2}, \dots, -(n(n-1))^{-1/2})$  and  $(2^{-1/2}, -(2^{-1/2}), 0, \dots, 0)$ . The computational cost is of order  $n^2$ .

Things are more cumbersome for  $w$ . In terms of computational time, the best family of processes that provide full identification of  $w$  is that of equally probable permutations of vectors  $(x, y, z, t, u, \dots, u)$  under the constraints  $x + y + z + t + (n-4)u = 0$  and  $x^2 + y^2 + z^2 + t^2 + (n-4)u^2 = 1$ . The computational cost is of order  $n^4$ . Since  $w$  has 21 components, we may compute  $E[T^2(k, \eta)|X, y_0]$  by permutations of seven different such vectors for three different values of  $k$  (say,  $k \in \{1, 2, 3\}$ ).

For instance, a typical post-war quarterly data set requires approximately 112 billions of computations of the statistic  $Z(k)$  for such processes. This is admittedly large, but remains manageable with current computers.<sup>22</sup>

As for the optimization, notice when  $E_\eta[\eta_1^4]$  is fixed all of the constraints are linear functions of  $E_\eta[\eta_1^6]$ ,  $E_\eta[\eta_1^8]$ ,  $E_\eta[\eta_1^3\eta_2^3]$ ,  $E_\eta[\eta_1^5\eta_2^3]$ ,  $E_\eta[\eta_1^4\eta_2^4]$ . The optimization problem for any given value of  $k$  and  $E_\eta[\eta_1^4]$  may be found using a simplex algorithm. A grid search over possible values of  $E_\eta[\eta_1^4]$  then solves the optimization problem. This optimization step is very quick as the simplex is of low dimension.

$$\begin{aligned} E_\eta[\eta_1^4\eta_2^4] &\geq (E_\eta[\eta_1^2\eta_2^2])^2 = \left(\frac{1-nE_\eta[\eta_1^4]}{n(n-1)}\right)^2 \\ \frac{E_\eta[\eta_1^6]-E_\eta[\eta_1^8]}{n-1} &= E_\eta[\eta_1^6\eta_2^2] \geq (E_\eta[\eta_1^3\eta_2])^2 = \left(\frac{E_\eta[\eta_1^4]}{n}\right)^2 \\ \frac{(1-n)E_\eta[\eta_1^4\eta_2^4]+E_\eta[\eta_1^4]-2E_\eta[\eta_1^6]+E_\eta[\eta_1^8]}{(n-1)(n-2)} &= E_\eta[\eta_1^4\eta_2^2\eta_3^2] \geq (E_\eta[\eta_1^2\eta_2\eta_3])^2 = \left(\frac{2nE_\eta[\eta_1^4]-1}{n(n-1)(n-2)}\right)^2 \\ \frac{3n(n-1)E_\eta[\eta_1^4\eta_2^4]-n(6E_\eta[\eta_1^4]-8E_\eta[\eta_1^6]+3E_\eta[\eta_1^8])+1}{n(n-1)(n-2)(n-3)} &= E_\eta[\eta_1^2\eta_2^2\eta_3^2\eta_4^2] \geq (E_\eta[\eta_1\eta_2\eta_3\eta_4])^2 \end{aligned}$$

and

$$(E_\eta[\eta_1\eta_2\eta_3\eta_4])^2 = 9 \left(\frac{1-2nE_\eta[\eta_1^4]}{n(n-1)(n-2)(n-3)}\right)^2.$$

Also  $\sum_{i=1}^n \eta_i^2 = 1$  allows us to formulate several other inequalities. For instance, all components of  $\eta$  are smaller than one which entails

$$\begin{aligned} \frac{E_\eta[\eta_1^6]-E_\eta[\eta_1^8]}{n-1} &= E_\eta[\eta_1^6\eta_2^2] \leq E_\eta[\eta_1^6] \\ \frac{(1-n)E_\eta[\eta_1^4\eta_2^4]-2E_\eta[\eta_1^6]+E_\eta[\eta_1^8]}{(n-1)(n-2)} &= E_\eta[\eta_1^4\eta_2^2\eta_3^2] \leq E_\eta[\eta_1^4\eta_2^2] = \frac{E_\eta[\eta_1^4]-E_\eta[\eta_1^6]}{n-1} \\ \frac{3n(n-1)E_\eta[\eta_1^4\eta_2^4]-n(6E_\eta[\eta_1^4]-8E_\eta[\eta_1^6]+3E_\eta[\eta_1^8])+1}{n(n-1)(n-2)(n-3)} &= E_\eta[\eta_1^2\eta_2^2\eta_3^2\eta_4^2] \leq E_\eta[\eta_1^2\eta_2^2\eta_3^2] \end{aligned}$$

and

$$E_\eta[\eta_1^2\eta_2^2\eta_3^2] = \frac{-3nE_\eta[\eta_1^4] + 2nE_\eta[\eta_1^6] + 1}{n(n-1)(n-2)}$$

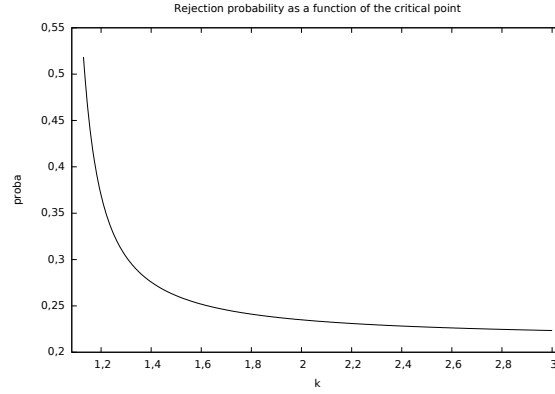
<sup>22</sup>The computation in this Section have been made with a Hewlett Packard Proliant DL 380-G7. The R program we wrote takes about 7.5 hours for a sample size  $n = 100$  after parallelization on 8 core Xenon processors. We are indebted to Stphane Nou for computational assistance. A  $n = 300$  sample requires one month on the same environment.



### 3.4 Illustration : AR(1) with linear trend

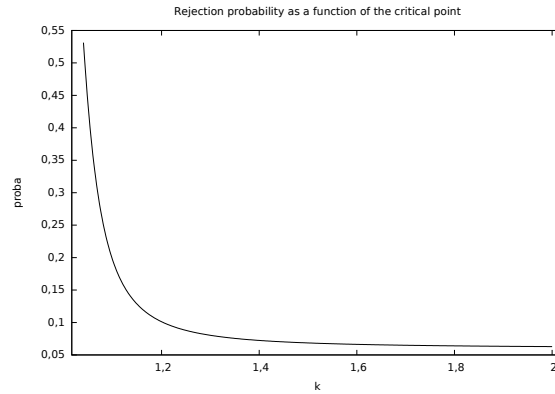
To illustrate the above procedure we consider the null hypothesis  $H_0 : \lambda(\bar{\mathbf{P}}) = 1$ . This case is of particular importance in finance and macro-econometrics. The random walk hypothesis is also famous to raise specific statistical difficulties. In particular, the asymptotic distributions depend on nuisance parameters for instance if the model incorporates deterministic trends. The above approach may be used to propose an exact unit-root test in a fully semi-parametric framework with possibly nuisance parameters such as deterministic trends (notice in particular the approach by Dufour and Roy [1985] does not allow for deterministic trends). Finally notice that in this example the explanatory variable depends only on the sample size. In such a case, the test may be tabulated once for all and the computational cost is a fixed one.

The following chart displays the bound on the rejection probability as a function of the critical point for a sample size equal to 100.



Again, we have a limit lower bound on the probability of rejection. For instance, with such a small sample size, it is impossible to reach the usual 5% level using our exact approach.

Things are a little better when we test for the presence of a lagged effect. Consider now the null hypothesis  $H_0 : \lambda(\bar{\mathbf{P}}) = 0$  we obtain the following chart:



In this case, we are very close to reach the usual 5%-level (the minimal bound we obtain is slightly below 6%).

## 4 Multivariate models

The test procedures derived so far use the exchangeability of the vector of “disturbances”. This property is well suited if observations may be considered as coming from interchangeable “cells” of information (for instance anonymous households, or dealers). It may however not be very convenient if blocks of observations are related to the same “cell” as in multivariate models.

Consider for instance the case where agent’s  $i$  consumption of coffee ( $C_i$ ) and tobacco ( $T_i$ ) are explained by his/her revenue  $R_i$ . In this case, conditional exchangeability may hold between two different individuals, but the consumption behaviors for the same individual are likely not exchangeable.

### 4.1 Line-exchangeable random matrices

We now consider a family  $\mathcal{M}_p$  of probability distributions for  $(Y, X)$ , where  $Y$ ,  $X$  are real random matrices of size  $n \times p$  and  $n \times K$  respectively (with  $K < n$ ). In this subsection, we consider the following assumption.

**Assumption 4** *For any  $\mathbf{P} \in \mathcal{M}_p$  there exist a single –non stochastic– matrix with  $K$  lines and  $p$  columns denoted  $\mathbf{B}(\mathbf{P})$  such that given  $X$  the conditional distribution the matrix  $Y - X\mathbf{B}(\mathbf{P})$  is invariant after any permutation of its lines and for all non-random vector  $\nu \neq 0$  of size  $p$  the random variable  $Y\nu - X\mathbf{B}(\mathbf{P})\nu$  is non degenerated.*

In what follows, ‘equation’ refers to observations corresponding to the same columns and ‘individual’ when we consider the same line. Moreover, in accordance with previous terminology, ‘disturbance’ now refers to the matrix  $Y - X\mathbf{B}(\mathbf{P})$ .

A comment on the intercept is again needed. As in Section 2, the above assumption (implicitly) implies that the matrix of parameters  $B$  does not contain an intercept in any equation. In particular, the method we propose below cannot be applied to test for equality of the intercept in two different equations.

It should be clear that the multivariate setting under study is a generalization of the framework studied in Section 2. Indeed, the distribution of

$$\text{vec}(Y) - \text{vec}(I_p \otimes X)\text{vec}(\mathbf{B}(\mathbf{P}))$$

is now invariant wrt to some permutations (instead of all permutations as in Section 2). Of course this generalization is closely linked to the fact that the correlations of the disturbances corresponding to two observations associated with the same individual are not assumed to equal the correlation of the disturbances corresponding to two observations associated with the same equation.

The following proposition further emphasizes this fact.

**Proposition 6** *Let  $U = [u_{i,j}]_{\substack{i \in \{1, \dots, n\} \\ j \in \{1, \dots, p\}}}$  be a random matrix. If for all*

permutation  $\sigma$  of  $\{1, \dots, n\}$  the vectors

$$(u_{1,1}; \dots; u_{n,1}; u_{1,2}; \dots; u_{n,2}; \dots, u_{1,p}; \dots, u_{n,p})$$

and

$$(u_{\sigma(1),1}; \dots, u_{\sigma(n),1}; u_{\sigma(1),2}; \dots; u_{\sigma(n),2}; \dots, u_{\sigma(1),p}; \dots; u_{\sigma(n),p})$$

share the same distribution we say that  $U$  is line-exchangeable.

Let  $M$  be any non-stochastic matrix of size  $p \times m$ . If the matrix  $U$  is line-exchangeable, then the matrix  $UM$  is line-exchangeable.

*Proof:* Let  $\sigma$  be a given permutation of  $\{1, \dots, n\}$  and define  $\Sigma$  the  $n \times n$  matrix such that  $\Sigma_{i,j} = 1$  if and only if  $\sigma(i) = j$  and  $\Sigma_{i,j} = 0$  otherwise. The matrix obtained from  $U$  after permutation of its lines according to  $\sigma$  is  $\Sigma U$ . Now we have

$$\Sigma(UM) = (\Sigma U)M.$$

Since  $U$  is line-exchangeable,  $(\Sigma U) \times M$  and  $U \times M$  have the same distribution, or equivalently  $\Sigma \times (UM)$  and  $U \times M$  have the same distribution, i.e.  $UM$  is line-exchangeable. Q.E.D

The above proposition implies as a corollary

**Corollary 2** *If the distribution of  $(Y, X)$  belongs to  $\mathcal{M}_p$ , then for any non stochastic matrix  $M$  of size  $p \times m$ , the distribution of  $(YM, X)$  belongs to  $\mathcal{M}_m$ .*

## 4.2 Testing $RBC = \Gamma_0$ with Wilk's ratio

We now consider the following null hypothesis  $H_0^{\mathcal{M}} : RBC = \Gamma_0$  where  $R$  is a given  $q \times K$  matrix with  $\text{rank}(R) = q$  and  $C$  is a given  $p \times r$  matrix with  $\text{rank}(C) = r$ . Form the above proposition, it is clear that we may consider the case  $C = I_p$  by post-multiplication of the matrices  $Y$  by  $C$  and the reparametrization  $\mathbf{B}C$ .

In this case, we shall now consider one of the equivalent to Fisher statistic in multivariate setting, the Wilk's ratio. Dufour and Khalaf [2002] show that under the null hypothesis this statistic is distributed as

$$\det(C'(Y - X\bar{\mathbf{B}})'P(Y - X\bar{\mathbf{B}})C)/\det(C'(Y - X\bar{\mathbf{B}})'P_0(Y - X\bar{\mathbf{B}})C)$$

where  $P$  and  $P_0$  are explicit projections matrices that depend on  $X$  and  $R$  (see Dufour and Khalaf [2002] for details) and  $\bar{\mathbf{B}}$  stands for the value of parameter  $\mathbf{B}$  associated with the DGP. Clearly, the above expression is invariant after normalizing  $(Y - X\bar{\mathbf{B}})$ . Also, similarly as in Section 2 we have

$$P(\iota_n \otimes \iota'_p) = P_0(\iota_n \otimes \iota'_p) = 0$$

as  $P$  and  $P_0$  are projection matrices. Hence the distribution of the Wilk's ratio under the null hypothesis is the same as that of

$$\det(H'PH)/\det(H'P_0H)$$

where  $H$  is a line-exchangeable matrix of size  $n \times c$  such that  $\iota'_n H = 0$  and  $\text{vec}(H)'H = 1$ .

As a straightforward consequence of corollary 2 above, in the case  $c = 1$  we may use the same bounds as in Section 2. Also remark that in this case, several usual multivariate statistics as the Lawley-Hotelling or Bartlett-Nanda-Pilai trace criteria all lead to the same type of decision rule. The practical importance of this particular case is extensively covered by Steward [1997].

To control the rejection probability of the event

$$\det(H'PH)/\det(H'P_0H) > k \Leftrightarrow M(k, H) = \det(H'PH) - k\det(H'P_0H) > 0$$

in the general case (that is if  $c > 1$ ) we may proceed as in the previous subsection. More precisely, we have to compute  $E_H[M(k, H)|X]$  and  $E_H[M(k, H)^2|X]$  as function of nuisance parameters related to multivariate moments of the elements of some line-exchangeable random matrix. Then for all  $k$  large enough so that we can insure  $E[M(k)] < 0$  whatever the distribution of  $H$  we may apply the Cantelli inequality. As the principle is exactly the same as in the previous section, we shall now sketch the procedure in the particular case  $c = 2$ . The approach for larger values of  $c$  follows along the same lines.

Block-decomposition leads to

$$H'PH = h'_1 P_{11} h_1 h'_2 P_{22} h_2 + (h'_1 P_{12} h_2)^2$$

where  $h_1$  and  $h_2$  are exchangeable vectors such that  $\sum_{i=1}^n h_{1,i} = \sum_{i=1}^n h_{2,i} = 0$  and

$$1 = \sum_{i=1}^n h_{1,i}^2 + \sum_{i=1}^n h_{2,i}^2$$

Proceeding as in Sections 2 and 3, we deduce that there exists two constants  $a, a'$  depending only on the matrices  $P$  and  $P_0$  such that

$$E_H[M(k, H)|X] = aE_H[h_{1,1}^2 h_{2,1}^2] + a'E_H[h_{1,1}^2 h_{2,2}^2] \quad (6)$$

and 7 constants (also depending solely on the matrices  $P$  and  $P_0$ ) such that

$$\begin{aligned} E_H[M(k, H)|X]^2 &= bE_H[h_{1,1}^4 h_{2,1}^4] + b'E_H[h_{1,1}^4 h_{2,1}^4] \\ &\quad + cE_H[h_{1,1}^3 h_{1,2}^3 h_{2,1}^3] + c'E_H[h_{1,1}^3 h_{1,2}^3 h_{2,3}^3] \\ &\quad + d_1 E_H[h_{1,1}^2 h_{1,2}^2 h_{2,1}^2 h_{2,2}^2] + d_2 E_H[h_{1,1}^2 h_{1,2}^2 h_{2,1}^2 h_{2,3}^2] \\ &\quad + d_3 E_H[h_{1,1}^2 h_{1,2}^2 h_{2,3}^2 h_{2,4}^2] \end{aligned} \quad (7)$$

Again, the relevant moments of monomials in  $h_{i,j}$  satisfy some constraints that may be used to derive a non trivial upper bound for the rejection probability under the null. For instance we get

$$\begin{aligned} E_H[h_{1,1}^2] &= E_H[h_{1,1}^2 \sum_{i=1}^n (h_{1,i}^2 + h_{2,i}^2)] \\ &= E_H[h_{1,1}^4] + (n-1)E_H[h_{1,1}^2 h_{1,2}^2] + E_H[h_{1,1}^2 h_{2,1}^2] + (n-1)E_H[h_{1,1}^2 h_{2,2}^2] \end{aligned}$$

Also from  $\sum_{i=1}^n (h_{1,i}^2 + h_{2,i}^2) = 1$  and the exchangeability of  $h_1$  and  $h_2$  we derive

$$E_H[h_{j,1}^2] \leq 1/n \quad \forall j \in \{1, 2\}.$$

Finally the Cauchy Schwarz inequality together with exchangeability imply inequalities that constrains quantities  $E[h_{1,1}^4 h_{2,1}^4], E[h_{1,1}^4 h_{2,2}^4]$  as in Section 3.2 above.

As for the practical implementation, specific distributions should again be used to derive the constants that appears in equations (6) and (7). The computational cost is similar to that of Section 3. In particular, using the same approaches as in previous sections, the presence of the term  $E_H[h_{1,1}^2 h_{1,2}^2 h_{2,3}^2 h_{2,4}^2]$  leads to a computational cost that increases as  $n^4$ .

## 5 Conclusion

We show that exact and powerful testing about the slope is possible for semi-parametric linear models. Moreover, the procedures proposed in this paper are easy to implement and rely on ordinary Fisher-type tests. A procedure derived from the Markov bound may be used to find “universal” upper bound to p-values even when direct access to the data is not available. Although this bound may be shown to be tight for some particular distribution of the exogenous variable, it is in general very conservative. To circumvent this problem improvements of the bound are proposed. These improvements involve higher moments. Two extensions are considered. First, in the AR(1) model with exogenous variables we show how to perform exact tests using a setting described in Dufour and Kiviet [1998]. Second, the results are extended to multivariate model. This lead us to introduce the notion of line-exchangeability of a random matrix. We prove that single linear constraints may be tested using Fisher test in our setting. For genuinely multivariate linear constraints a method to control the rejection probability of a procedure based on Wilk’s ratio is sketched.

Several other subjects of interest related to this work are planned to be studied in future works. First linear inference with instruments and/or simultaneous models may also been studied along the lines sketched in Section 4. Also, although exact, non trivial testing about the slope with heteroskedasticity of unknown form is likely not feasible in our setup, further results may be derived for specific forms of heteroskedasticity.

From a more mathematical standpoint, deriving tight inequalities for higher moments raises interesting problems. It has direct impact on the balance between the –potential– improvement provided by higher moment versions and the extra difficulties created by the control of unknown nuisance parameters.

Finally, deriving the more efficient programs to compute the decomposition of the moments of the statistics on the relevant sequence of multivariate high–order moments of self–normalized, self–centered processes is a challenging computational task.

## References

- Aldous, D. [1983] “Exchangeability and related topics”, In *Ecole d Eté de Probabilités de SaintFlour*, **XIII**, 1198, Springer-Verlag, Heidelberg. Lecture Notes in Mathematics 1117.
- M.C. Beaulieu, J.M Dufour, L. Khalaf [2013] “Exact confidence sets and goodness-of-fit test methods for stable distributions”, *Journal of Econometrics*, forthcoming.
- Bahadur and Savage [1956] “The nonexistence of certain statistical procedures in nonparametric problems”, *Annals of Mathematical Statistics*, **27**, 1115-

1122.

J. Dedecker and F. Merlevède [2002] ‘Necessary and Sufficient Conditions for the Conditional Central Limit Theorem’ *The Annals of Probability*, **30** 1044-1081.

J.M. Dufour [2003] ”Identification, Weak Instruments and Statistical Inference in Econometrics. *Canadian Journal of Economics*, **36**, 767-808.

J.M. Dufour [1997] ”Some Impossibility Theorems in Econometrics with Applications to Structural and Dynamic Models”, *Econometrica*, **65**, 1997, 1365-1388.

J.M. Dufour [2006] “Monte Carlo Tests with nuisance parameters: A general approach to finite-sample inference and nonstandard asymptotics in econometrics”, *Journal of Econometrics*, **133**, 443-477.

J.M. Dufour, F. Jouneau-Sion and O. Torrès [2008] “Testability and methods of moments in nonparametric and semiparametric models”, mimeo.

J.M. Dufour and L. Khalaf [2002] “Simulation-Based Finite and Large Sample Tests in Multivariate Regressions”, *Journal of Econometrics*, **111**, 303-322.

J.M. Dufour and J. Kiviet [1998] “Exact inference methods for first-order autoregressive distributed lag models”, *Econometrica*, **66**, 79-104.

J.M. Dufour and R. Roy [1985] “Some Robust Exact Results on Sample Autocorrelations and Tests of Randomness”, *Journal of Econometrics*, **29**, 257-273.

J.M. Dufour and R. Roy [1983] “Some Robust Exact Results on Sample Autocorrelations and Tests of Randomness”, *Journal of Econometrics*, **29**, 257-273.

J.M. Dufour and M. Taamouti [2005] “Projection-based statistical inference in linear structural models with possibly weak instruments” *Econometrica*, **73**, 1351-1365.

X. Gabaix [2009] “Power Laws in Economics and Finance *Annual Review of Economics*, **1**, 255-293

J. Galambos [1982], “Exchangeability” in *Encyclopedia of Social Sciences*, **2**, 573-577, Wiley, New-York.

O. Gossner and K. Schlag [2013] “Finite-sample exact tests for linear regressions with bounded dependent variables”, *Journal of econometrics*, **177**, 75-84.

D. Kalman [2000] “A Matrix Proof of Newton’s Identities”, *Mathematics Magazine*, **73**, 313-315

M. Kendall, A. Stuart, K. Ord [1983], *Advanced Theory of Statistics*, **3**, Wiley.

Kirch C. [2007] Block permutation principles for the change analysis of dependent data *Journal of Statistical Planning and Inference* **137**, 2453-2474.

P. Mc Cullagh [2005], “Exchangeability and regression models”, in *Celebrating Statistics (Festschrift for D.R. Cox)*, 89-113.

A. Mikusheva [2007] “Uniform inference in autoregressive models”, *Econometrica*, **75**, 1411-1452.

J. Nankervis and N.E. Savin [1985] “Testing the autoregressive parameter with the  $t$ -statistic” *Journal of Econometrics*, **27**, 143-161.

M.B. Nathanson [2000] *Elementary Methods in Number Theory*, Graduate Texts in Mathematics 195. Springer-Verlag.

B. Pötscher, [2009] “Confidence Sets Based on Sparse Estimators Are Necessarily Large”, *Sankhya*, **71-A**, 1-18.

K. G. Steward [1997] “ Exact Testing in Multivariate Regression”, *Econometric Reviews*, **16**, 321–352.